GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 16, 2003, 12:13:01 ; Search time 32 Seconds
(without alignments)
766.238 Million cell updates/sec

Title: US-10-015-967-2
Perfect score: 644
Sequence: 1 MKVLISSLLLLLPLMLMSMV.....SRACQQFLKQCQLRSFALPL 119

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	322	50.0	128	11	Q8R3U6	Q8r3u6 mus musculu
2	77.5	12.0	477	5	O62022	O62022 caenorhabdi
3	76.5	11.9	411	13	Q9W745	Q9w745 xenopus lae
4	72.5	11.3	3313	11	O88278	O88278 rattus norv
5	71.5	11.1	108	10	O82328	O82328 arabidopsis
6	71	11.0	188	4	Q8TEV2	Q8tev2 homo sapien
7	71	11.0	3190	5	O01368	O01368 drosophila
8	71	11.0	3275	5	Q9W321	Q9w321 drosophila
9	70.5	10.9	160	10	Q9SQH1	Q9sqh1 arachis hyp
10	70.5	10.9	1290	3	Q96UA3	Q96ua3 neurospora
11	70	10.9	117	4	Q9BRN8	Q9brn8 homo sapien
12	70	10.9	743	5	O97120	O97120 schistosoma
13	70	10.9	1095	5	Q21852	Q21852 caenorhabdi
14	70	10.9	1664	5	Q9TVQ2	Q9tvq2 caenorhabdi
15	69.5	10.8	221	12	Q85299	Q85299 orf virus.
16	69.5	10.8	468	4	Q9UC32	Q9uc32 homo sapien
17	69	10.7	111	10	Q9FWV9	Q9fwv9 oryza sativ
18	69	10.7	200	16	Q9HVU8	Q9hvu8 pseudomonas
19	69	10.7	271	4	Q9HBS0	Q9hbs0 homo sapien
20	69	10.7	378	3	O74213	O74213 aspergillus
21	69	10.7	413	10	Q9C8W2	Q9c8w2 arabidopsis
22	69	10.7	454	13	Q90ZE8	Q90ze8 oncorhynchu
23	69	10.7	527	10	Q9SKW4	Q9skw4 arabidopsis
24	69	10.7	1081	6	Q9BE73	Q9be73 macaca fasc
25	68.5	10.6	381	5	Q26641	Q26641 strongyloce
26	68.5	10.6	405	13	Q9PS79	Q9ps79 xenopus. xc
27	68.5	10.6	422	11	Q922X0	Q922x0 mus musculu
28	68.5	10.6	427	6	Q95MH5	Q95mh5 saguinus oe
29	68.5	10.6	1006	16	P73012	P73012 synechocyst
30	68	10.6	189	14	Q99IU7	Q99iu7 uncultured
31	68	10.6	283	10	Q43508	Q43508 lycopersico
32	68	10.6	291	13	Q9IAK3	Q9iak3 xenopus lae
33	68	10.6	539	10	Q9XHV3	Q9xhv3 oryza sativ
34	67.5	10.5	453	13	Q9PTN2	Q9ptn2 brachydanio
35	67.5	10.5	1069	5	Q9BPS2	Q9bps2 bombyx mori
36	67	10.4	376	10	Q9LXW3	Q9lxw3 arabidopsis
37	67	10.4	454	13	Q8QGE2	Q8qge2 salmo salar
38	67	10.4	743	12	Q83111	Q83111 mouse adeno
39	67	10.4	1062	11	O88849	O88849 mus musculu
40	67	10.4	1062	11	O89019	O89019 mus musculu
41	66.5	10.3	103	17	Q9YCZ2	Q9ycz2 aeropyrum p
42	66.5	10.3	419	11	Q62681	Q62681 rattus norv
43	66.5	10.3	422	13	Q91430	Q91430 brachydanio
44	66.5	10.3	2571	11	Q8R4Y4	Q8r4y4 mus musculu
45	66.5	10.3	3301	11	Q91ZI0	Q91zi0 mus musculu

ALIGNMENTS

RESULT 1
Q8R3U6
ID Q8R3U6 PRELIMINARY; PRT; 128 AA.
AC Q8R3U6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to hypothetical gene LOC125944.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC024561; AAH24561.1; -.
SQ SEQUENCE 128 AA; 14043 MW; CCE67DA04D23D7DF CRC64;

Query Match 50.0%; Score 322; DB 11; Length 128;
Best Local Similarity 70.9%; Pred. No. 2.1e-29;
Matches 61; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

```
Qy      1 MKVLISSLLLLLPLMLMSMVSSSLNPGVARGHRDRGQASRRWLQEGGQECECKDWFLRAP 60
          ||:| |  |||||:|||||| || |||||| | |:  | |||| ||||||||||||:||
Db      1 MKLLASPFLLLLPVMLMSMVFSSPNPGVARSHGDQHLAPRRWLLEGGQECECKDWFLQAP 60

Qy     61 RRKFMTVSGLPKKQCPCDHFKGNVKK 86
          :||   | | |:||||||| ||  ||
Db     61 KRKATAVLGPPRKQCPCDHVKGREKK 86
```

RESULT 2
O62022
ID O62022 PRELIMINARY; PRT; 477 AA.
AC O62022; O18293;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

```
DE   ZK131.11 protein.
GN   ZK131.11.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Percy C.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z93373; CAB07552.1; -.
DR   EMBL; Z83245; CAB07552.1; JOINED.
DR   EMBL; Z83245; CAB05840.1; -.
DR   EMBL; Z93373; CAB05840.1; JOINED.
SQ   SEQUENCE   477 AA;  53385 MW;  86F5257F9EB6DD90 CRC64;

  Query Match             12.0%;  Score 77.5;  DB 5;  Length 477;
  Best Local Similarity   34.7%;  Pred. No. 1.2;
  Matches   25;  Conservative    6;  Mismatches   32;  Indels    9;  Gaps    3;

Qy      46 GGQECECKDWFLRAPRR---KFMTVSGLPKKQ--CPCDHFKGNVKKTRHQRHHRKPNKHS 100
           ||   |    |||| :|   |  | : |||||     |:|: :        ||  ||
Db     290 GGPNAESNTTFLRATKRFFKKIYTSATLPKKQNSTSMDNFQKSSVFFEDTAHHHNPNMDG 349

Qy     101 RACQQFLKQCQL 112
           |     ||:  |
Db     350 R----ILKKAPL 357


RESULT 3
Q9W745
ID   Q9W745      PRELIMINARY;      PRT;   411 AA.
AC   Q9W745;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   COUP transcription factor 1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Langlois M.C., Vanacker J.M., Holland N.D., Escriva H., Queva C.,
RA   Laudet V., Holland L.Z.;
RT   "Amphicoup-TF, a nuclear orphan receptor of the lancelet Branchiostoma
RT   floridae, is implicated in retinoic acid signaling pathways.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF157558; AAD42224.1; -.
DR   HSSP; P19793; 2NLL.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   411 AA;  45467 MW;  950E285B7FC336FB CRC64;

  Query Match             11.9%;  Score 76.5;  DB 13;  Length 411;
  Best Local Similarity   26.0%;  Pred. No. 1.4;
  Matches   26;  Conservative   16;  Mismatches   25;  Indels   33;  Gaps    6;

Qy      26 PGVA--RGHRDRGQASRRWLQ-----------EGGQECE-CKDWFLRAPRRKFMTVSGL 70
           || |  :| :  ||: :: ::             |   || || :| |: ||  :| :
Db      53 PGAAGDKGQQGSGQSQQQHIECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRN-LTYTCR 111

Qy      71 PKKQCPCDHFKGNVKKTRHQRHHRKPNKHSRACQQFLKQC 110
           : || |            :|||   :: |     ||:|
Db     112 ANRNCPID----------QHHRNQCQYCR-----LKKC 134


RESULT 4
O88278
ID   O88278      PRELIMINARY;      PRT;  3313 AA.
AC   O88278;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   MEGF2.
GN   MEGF2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE=98360089; PubMed=9693030;
RA   Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT   "Identification of high-molecular-weight proteins with multiple EGF-
RT   like motifs by motif-trap screening.";
RL   Genomics 51:27-34(1998).
CC   -!- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.
DR   EMBL; AB011528; BAA32459.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR001879; hormn_receptor.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00028; cadherin; 9.
DR   Pfam; PF00008; EGF; 6.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 8.
DR   SMART; SM00180; EGF_Lam; 1.
DR   SMART; SM00001; EGF_like; 6.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR   PROSITE; PS00232; CADHERIN_1; 6.
DR   PROSITE; PS50268; CADHERIN_2; 9.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_6.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
KW   Calcium-binding; Cell adhesion; EGF-like domain; Glycoprotein.
SQ   SEQUENCE   3313 AA;  359348 MW;  B11DA09517288764 CRC64;

  Query Match             11.3%;  Score 72.5;  DB 11;  Length 3313;
  Best Local Similarity   27.6%;  Pred. No. 35;
  Matches   35;  Conservative    9;  Mismatches   40;  Indels   43;  Gaps    7;

Qy      24 LNPGVARGHRDRGQASRRWLQEG--GQECECKDWFL----------RAPRRKFMTVSGLP 71
           |||        : | | | || |   |  |:|   :           : ||  : : :  |
Db    1978 LNP-------CQNQGSCRHLQGGPHGYTCDCASGYFGQHCEHRMDQQCPRGWWGSPTCGP 2030

Qy      72 KKQCPCDHFKG---NVKKTRHQRH----HRKP--------------NKHSRACQQFLKQC 110
```

```
Db    2031 ---CNCDVHKGFDPNCNKTSGQCHCKEFHYRPRGSDSCLPCDCYPVGSTSRSCAPHSGQC 2087

Qy     111 QLRSFAL 117
Db    2088 PCRPGAL 2094


RESULT 5
O82328
ID   O82328      PRELIMINARY;      PRT;   108 AA.
AC   O82328;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GAST1/GASA-like protein.
GN   AT2G14900.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005396; AAC61287.1; -.
DR   InterPro; IPR003854; GASA.
DR   Pfam; PF02704; GASA; 1.
SQ   SEQUENCE   108 AA;  11381 MW;  93206EAE3BB95DF5 CRC64;

  Query Match             11.1%;  Score 71.5;  DB 10;  Length 108;
  Best Local Similarity   25.9%;  Pred. No. 1.3;
  Matches   29;  Conservative   18;  Mismatches   38;  Indels   27;  Gaps    7;

Qy       1 MKVLISSLLLL-LPLMLMSMVSSSLNPGVARG------HRDRGQASRRWLQEGGQECE-- 51
Db       1 MKIIVSILVLASLLLISSSLASATISDAFGSGAVAPAPQSKDGPALEKWC---GQKCEGR 57

Qy      52 CKDWFLRAPRRKFMTVSGLPKKQCPCDHFKGNVKKTRHQRHHRKPNKHSRAC 103
Db      58 CKE---AGMKDRCLKYCGICCKDCQC------VPSGTY------GNKHECAC 94


RESULT 6
Q8TEV2
ID   Q8TEV2      PRELIMINARY;      PRT;   188 AA.
AC   Q8TEV2;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Vascular endothelial growth factor B isoform VEGF-B167.
GN   VEGFB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96325041; PubMed=8702615;
RA   Olofsson B., Pajusola K., von Euler G., Chilov D., Alitalo K.,
RA   Eriksson U.;
RT   "Genomic organization of the mouse and human genes for vascular
RT   endothelial growth factor B (VEGF-B) and characterization of a second
RT   splice isoform.";
RL   J. Biol. Chem. 271:19310-19317(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98032228; PubMed=9365524;
RA   Joukov V., Kaipainen A., Jeltsch M., Pajusola K., Olofsson B.,
RA   Kumar V., Eriksson U., Alitalo K.;
RT   "Vascular endothelial growth factors VEGF-B and VEGF-C.";
RL   J. Cell. Physiol. 173:211-215(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Jeltsch M., Pajusola K., Chilov D., Alitalo K.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF468110; AAL79000.1; -.
SQ   SEQUENCE   188 AA;  21314 MW;  F04654D5A3626095 CRC64;

  Query Match             11.0%;  Score 71;  DB 4;  Length 188;
  Best Local Similarity   29.7%;  Pred. No. 2.7;
  Matches   22;  Conservative   13;  Mismatches   15;  Indels   24;  Gaps    6;

Qy      43 LQEGGQECECKDWFLRAPRRKFMTVS-GLPKKQCPCDHFKGNVKKTRHQRHHRKPNKHSR 101
Db     116 LEEHSQ-CECR------PKKKDSAVKPDSPRPLCP----------RCTQHHQRPD--PR 155

Qy     102 ACQQFLKQCQLRSF 115
Db     156 TCR---RRCRRRSF 166


RESULT 7
O01368
ID   O01368      PRELIMINARY;      PRT;  3190 AA.
AC   O01368;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CREB-binding protein homolog.
GN   NEJ OR CG15319.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97263578; PubMed=9109493;
RA   Akimaru H., Chen Y., Dai P., Hou D.X., Nonaka M., Smolik S.M.,
RA   Armstrong S., Goodman R.H., Ishii S.;
RT   "Drosophila CBP is a co-activator of cubitus interruptus in hedgehog
RT   signalling.";
RL   Nature 386:735-738(1997).
DR   EMBL; U88570; AAB53050.1; -.
DR   TRANSFAC; T03236; -.
DR   FlyBase; FBgn0015624; nej.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR003101; KIX.
DR   InterPro; IPR000197; TAZ_finger.
DR   InterPro; IPR000433; Znf_ZZ.
DR   Pfam; PF00439; bromodomain; 1.
DR   Pfam; PF02172; KIX; 1.
DR   Pfam; PF02135; zf-TAZ; 2.
DR   Pfam; PF00569; ZZ; 1.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00291; ZnF_ZZ; 1.
DR   PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
```

```
DR   PROSITE; PS01357; ZF_ZZ_1; UNKNOWN_1.
SQ   SEQUENCE   3190 AA;  331879 MW;  E53526F78BC055A8 CRC64;

  Query Match             11.0%;  Score 71;  DB 5;  Length 3190;
  Best Local Similarity   26.7%;  Pred. No. 50;
  Matches   23;  Conservative   11;  Mismatches   30;  Indels   22;  Gaps    3;

Qy      25 NPGVARGHRDRGQASRRWLQEGGQECECKDWFLRAPRRKFMTVSGLPKKQCPCDHFKGNV 84
           ||  |     | |: :| :|     |:|:|   | |               |   | | |
Db    2390 NPQEA-----RKQSIQRCIQSLAHACQCRDANCRLP--------------SCQKMKLVV 2429

Qy      85 KKTRHQRHHRKPNKHSRACQQFLKQC 110
           : |:: :  ||||     |:| :  |
Db    2430 QHTKNCK--RKPNGGCPICKQLIALC 2453


RESULT 8
Q9W321
ID   Q9W321      PRELIMINARY;      PRT;  3275 AA.
AC   Q9W321;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CG15319 protein.
GN   NEJ OR CG15319.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003448; AAF46516.1; -.
DR   FlyBase; FBgn0015624; nej.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR003101; KIX.
DR   InterPro; IPR000197; TAZ_finger.
DR   InterPro; IPR000433; Znf_ZZ.
DR   Pfam; PF00439; bromodomain; 1.
DR   Pfam; PF02172; KIX; 1.
DR   Pfam; PF02135; zf-TAZ; 2.
DR   Pfam; PF00569; ZZ; 1.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00291; ZnF_ZZ; 1.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PROSITE; PS01357; ZF_ZZ_1; UNKNOWN_1.
SQ   SEQUENCE   3275 AA;  340672 MW;  E9944C3BFEC0E7AA CRC64;

  Query Match             11.0%;  Score 71;  DB 5;  Length 3275;
  Best Local Similarity   26.7%;  Pred. No. 51;
  Matches   23;  Conservative   11;  Mismatches   30;  Indels   22;  Gaps    3;

Qy      25 NPGVARGHRDRGQASRRWLQEGGQECECKDWFLRAPRRKFMTVSGLPKKQCPCDHFKGNV 84
           ||  |     | |: :| :|     |:|:|   | |               |   | | |
Db    2389 NPQEA-----RKQSIQRCIQSLAHACQCRDANCRLP--------------SCQKMKLVV 2428

Qy      85 KKTRHQRHHRKPNKHSRACQQFLKQC 110
           : |:: :  ||||     |:| :  |
Db    2429 QHTKNCK--RKPNGGCPICKQLIALC 2452


RESULT 9
Q9SQH1
ID   Q9SQH1      PRELIMINARY;      PRT;  160 AA.
AC   Q9SQH1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Allergen.
GN   ARA H 7.
OS   Arachis hypogaea (Peanut).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC   Arachis.
OX   NCBI_TaxID=3818;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VIRGINIA; TISSUE=SEED;
RX   MEDLINE=99406463; PubMed=10474031;
RA   Kleber-Janke T., Crameri R., Appenzeller U., Schlaak M., Becker W.M.;
RT   "Selective cloning of peanut allergens, including profilin and 2S
RT   albumins, by phage display technology.";
RL   Int. Arch. Allergy Immunol. 119:265-274(1999).
DR   EMBL; AF091737; AAD56719.1; -.
DR   InterPro; IPR003612; AAI.
DR   InterPro; IPR001768; Try/amyl_inhbtr.
DR   Pfam; PF00234; tryp_alpha_amyl; 1.
DR   SMART; SM00499; AAI; 1.
SQ   SEQUENCE   160 AA;  18417 MW;  9F9E4CEE68808D4C CRC64;

  Query Match             10.9%;  Score 70.5;  DB 10;  Length 160;
  Best Local Similarity   22.1%;  Pred. No. 2.6;
  Matches   33;  Conservative   19;  Mismatches   46;  Indels   51;  Gaps    6;

Qy       1 MKVLISSLLLLLPLMLMSMVSSSLNPGVARGHRDRGQASRRWLQEGGQECECKDWFLRAP 60
           | | :| |: ||  :|:   ::  :|       |||    ||      : :|:    ||
Db       1 MMVKLSILVALLGALLVVASATRWDP-------DRGSRGSRWDAPSRGDDQCQRQLQRAN 53

Qy      61 RRKFMTVSGLPKKQCPC-DHFKGNVKKTRHQRHHRKP---------------NKHSRAC 103
             |             || :| :  |:: : |     |               |:  | |
Db      54 LR-------------PCEEHMRRRVEQEQEQEQDEYPYSRRGSRGRQPGESDENQEQRCC 100

Qy     104 Q-------------QFLKQ-CQLRSFALP 118
```

```
                    | |:|  | :|| :|
Db    101 NELNRFQNNQRCMCQALQQILQNQSFWVP 129


RESULT 10
Q96UA3
ID   Q96UA3      PRELIMINARY;      PRT;  1290 AA.
AC   Q96UA3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 137.8 kDa protein.
GN   B11B22.060.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL356834; CAD11425.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1290 AA;  137768 MW;  1076A879259BEB70 CRC64;

  Query Match             10.9%;  Score 70.5;  DB 3;  Length 1290;
  Best Local Similarity   24.5%;  Pred. No. 22;
  Matches   26;  Conservative   18;  Mismatches   31;  Indels   31;  Gaps    6;

Qy     33 RDRGQASRRWLQEGGQECECKDWFLRAPRRKFMTVS---GLPKKQCPCDHFKGNVKK--T 87
          |::    ::| ||| :  :     ||||      ||   | | :  |   ::|:::
Db    454 RNKKMKKQKWKQEGSKGSK-----LRAPESAMPPVSVVWGSPSRAWPLPRWQGHIRNGDV 508

Qy     88 RHQRHHR--------------KPNKHSRACQQF--LKQCQLRSFA 116
           || |::               |   |    |||  |:: :|  :|
Db    509 GHQSHNQHMYMDEGYVRRLWSEKLEPH----QQFARLRKAELALYA 550


RESULT 11
Q9BRN8
ID   Q9BRN8      PRELIMINARY;      PRT;   117 AA.
AC   Q9BRN8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 13.1 kDa protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EYE;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC006151; AAH06151.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   117 AA;  13085 MW;  77A87C1DC40665DC CRC64;

  Query Match             10.9%;  Score 70;  DB 4;  Length 117;
  Best Local Similarity   29.5%;  Pred. No. 2.1;
  Matches   26;  Conservative    7;  Mismatches   23;  Indels   32;  Gaps    5;

Qy     16 LMSMVSSSLNPGVAR---GHRDRGQASRRWLQEGGQECECKDWFLRAPRRKFMTVSGLPK 72
          |:|||     || ||      |  | :| ||               |||::  :||||:
Db     18 LLSMV-----PGPARPPGSCWDPTQCTRTWL------------LSHTPRRRW--ISGLPR 58

Qy     73 KQCP----------CDHFKGNVKKTRHQ 90
```

```
          |          ||   |     ||:
Db     59 ASCRLGEEPPPLPYCDQAYGEELSIRHR 86


RESULT 12
O97120
ID   O97120      PRELIMINARY;      PRT;   743 AA.
AC   O97120;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Retinoic acid receptor RXR.
OS   Schistosoma mansoni (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC   Schistosomatoidea; Schistosomatidae; Schistosoma.
OX   NCBI_TaxID=6183;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99143112; PubMed=9988692;
RA   Freebern W.J., Osman A., Niles E.G., Christen L., LoVerde P.T.;
RT   "Identification of a cDNA encoding a retinoid X receptor homologue
RT   from Schistosoma mansoni. Evidence for a role in female-specific gene
RT   expression.";
RL   J. Biol. Chem. 274:4577-4585(1999).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF094759; AAD16119.1; -.
DR   HSSP; P19793; 1RXR.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   743 AA;  82558 MW;  4CA9D287270B64DA CRC64;

  Query Match             10.9%;  Score 70;  DB 5;  Length 743;
  Best Local Similarity   28.4%;  Pred. No. 14;
  Matches   29;  Conservative   15;  Mismatches   28;  Indels   30;  Gaps    7;

Qy     18 SMVSSSLNP-GVARGHRDRGQASRRWLQEGGQECE-CKDWFLRAPRRKFMTV---SGLPK 72
          |: :::|||  |  | :  |:        |   || || :| |  |:: : |   ||
Db    261 SVNTTNLNPICVICGDKASGK------HYGVISCEGCKGFFKRTVRKQLVYVCRESG--- 311

Qy     73 KQCP--------CDH-------FKGNVKKTRHQRHHRKPNKH 99
           |||         | |         ||  |:   :  ||:|: :
Db    312 -QCPVDRRKRTRCQHCRFEQCLAKGMKKEAVQEERHRQPSSN 352


RESULT 13
Q21852
ID   Q21852      PRELIMINARY;      PRT;  1095 AA.
AC   Q21852;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   R09A8.1 protein.
GN   R09A8.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilkinson J.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z68009; CAA92003.1; -.
SQ   SEQUENCE   1095 AA;  124431 MW;  33DD93B1D906752C CRC64;

  Query Match             10.9%;  Score 70;  DB 5;  Length 1095;
  Best Local Similarity   31.6%;  Pred. No. 21;
  Matches   24;  Conservative   10;  Mismatches   32;  Indels   10;  Gaps    3;

Qy      20 VSSSLNPGVARGHRDRGQASRRWLQEGGQECECKDWFLRAPRRKFMTVSGLPKKQCPCDH 79
           | |||  |  |   ||  | :|:   |  | ||: |     :::        ||:   ||
Db     305 VQSSLMVGGNRSTDDRVNA-QRFASTGFVEKECRRWDQLVEKKQ--------KKEVNSDH 355

Qy      80 FKGNVKKTRHQRHHRK 95
            | | : | |  |: :
Db     356 KKAN-RITSHLEHNSR 370


RESULT 14
Q9TVQ2
ID   Q9TVQ2      PRELIMINARY;      PRT;  1664 AA.
AC   Q9TVQ2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Y64G10A.7 protein.
GN   Y64G10A.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mortimore B.J.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Ainscough R.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL117206; CAB60454.1; -.
DR   EMBL; AL110498; CAB60454.1; JOINED.
DR   EMBL; AL110498; CAB57911.1; -.
DR   EMBL; AL117206; CAB57911.1; JOINED.
DR   HSSP; P00736; 1APQ.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR002049; Laminin_EGF.
DR   Pfam; PF00008; EGF; 25.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00179; EGF_CA; 4.
DR   SMART; SM00001; EGF_like; 18.
DR   PROSITE; PS00010; ASX_HYDROXYL; 4.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_22.
DR   PROSITE; PS01186; EGF_2; 24.
DR   PROSITE; PS01187; EGF_CA; 3.
KW   Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ   SEQUENCE   1664 AA;  179279 MW;  A69F093B4C705832 CRC64;

  Query Match             10.9%;  Score 70;  DB 5;  Length 1664;
  Best Local Similarity   30.0%;  Pred. No. 33;
  Matches   21;  Conservative   10;  Mismatches   35;  Indels    4;  Gaps    3;

Qy      46 GGQECECKDWF-LRAPRRKFMTVSGLPKKQCPCDHFKGNVKKTRH--QRHHRKPNKHSRA 102
           |  :|:|   | |   ||  : :    |    |:||  ||| |     |   :  :  |
Db     273 GRAQCQCYPGFHLSYDRRSCVDIDECAKNN-GCEHFCENVKGTYRCKCREGYQLGRDGRT 331

Qy     103 CQQFLKQCQL 112
           |:: |  ||:
Db     332 CEEMLGGCQV 341


RESULT 15
Q85299
ID   Q85299      PRELIMINARY;      PRT;  221 AA.
AC   Q85299;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   Homologue of retroviral pseudoprotease.
OS   Orf virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Parapoxvirus.
OX   NCBI_TaxID=10258;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90266454; PubMed=2129563;
RA   Fraser M., Hill D.F., Mercer A.A., Robinson A.J.;
RT   "Sequence analysis of the inverted terminal repetition in the genome
RT   of the parapoxvirus, orf virus.";
RL   Virology 176:379-389(1990).
DR   EMBL; M30023; AAA46789.1; -.
SQ   SEQUENCE   221 AA;  25266 MW;  25659ABD386C2317 CRC64;

  Query Match             10.8%;  Score 69.5;  DB 12;  Length 221;
  Best Local Similarity   29.2%;  Pred. No. 4.7;
  Matches   33;  Conservative    8;  Mismatches   37;  Indels   35;  Gaps    6;

Qy      26 PGVARGHRDRGQASRRWLQEGGQECECKDWFLRAPRRKFMTVSGLPKK---------QCP 76
           |  ||| | ||  :||    ||     : |  || ||     || | :         | |
Db      83 PAAARGARRRGCCARRARGGGGWP---RRWRPRAARRG----SGPPARAPAAALAPDQAP 135

Qy      77 CDHFKGNVKKTR-------------HQR--HHRKPNKHSRACQQFLKQCQLRS 114
                :  ||| |               | |    |: ::  :||      |  ||
Db     136 ----RSKVKKDRLAVETLPPQPRTPHTRLPPARRQHRSQQACTPRRAGCSARS 184


Search completed: April 16, 2003, 12:16:13
Job time : 36 secs
```